Answer Key

Chapter 1 Review Packet

1. A statistician collects data for a study about cars. They look at a few different models of Toyotas (Highlander, Celica, Camry, Corolla, Avalon, Sequoia, 4Runner, Tundra, Tacoma, Rav4, Yaris, etc.) and they measure the following things on each car:

Miles per gallons Q
# of doors C
Weight Q
Starting price Q
Height Q
SUV or compact C
Size of engine (V4, V6, or V8) C
Average cost for car insurance Q

   a. What are the individuals?

   the different models of the car.

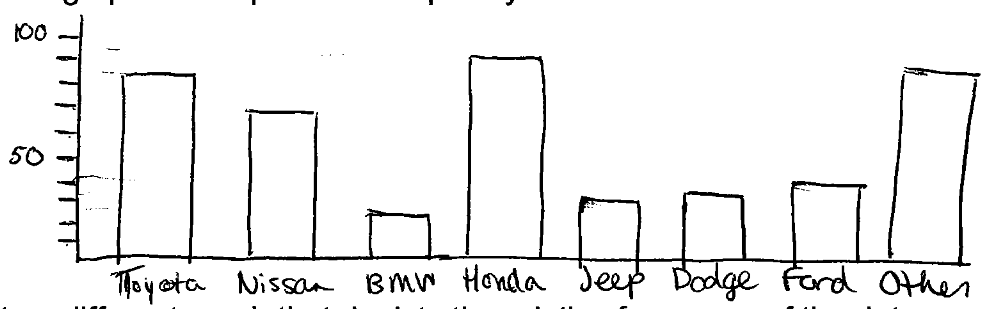
   b. What are the variables? Label each as Categorical or Quantitative.

2. There was a survey of the brands of cars that students at CB South drive. The results were as follows:

Brand	Count	Brand	Count
Toyota	85	Jeep	32
Nissan	71	Dodge	34
BMW	27	Ford	41
Honda	95	Other	90

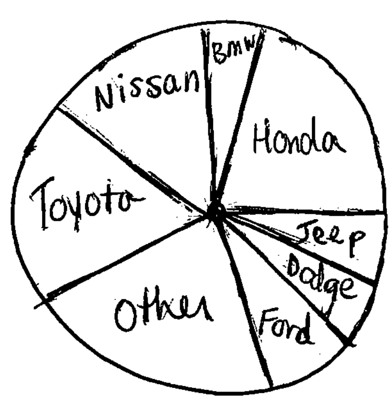
total: 475

   a. Create a graph that depicts the frequency of the data.



   b. Create a different graph that depicts the relative frequency of the data.

- Toyota 0.179
- Nissan 0.149
- BMW 0.057
- Honda 0.2
- Jeep 0.067
- Dodge 0.072
- Ford 0.086
- Other 0.189



3. Make a back-to-back split stemplot of the following data:

Reading Scores										
4th Graders	12	15	18	20	20	22	25	26	28	29
	31	32	35	35	35	36	37	39	40	42
7th Graders	1	12	15	18	18	20	23	23	24	25
	27	28	30	30	31	33	33	33	35	36

4th		7th
	0	1
	0	
2	1	2
85	1	588
200	2	0334
9865	2	578
21	3	001333
976555	3	56
20	4	
	4	

4. Make a comparison between 4th grade and 7th grade reading scores based on your stemplots.

7th - roughly symmetric.
lower range compared to 4th
center ~ 20's

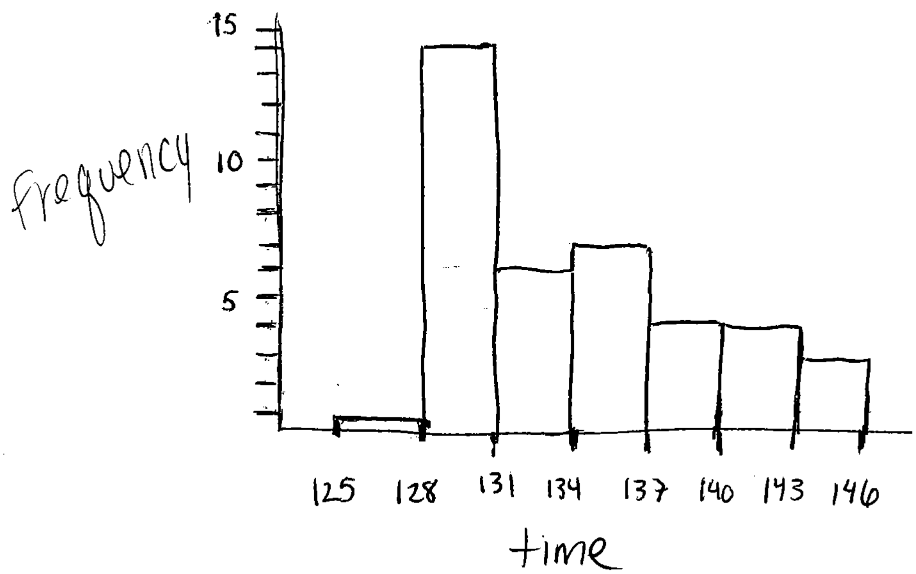
4th - higher range
center ~ 30's
slight left skew

5. The following data gives the times (in minutes, rounded to the nearest minute) for the winning man in the Boston Marathon in the years 1959 to 1997.

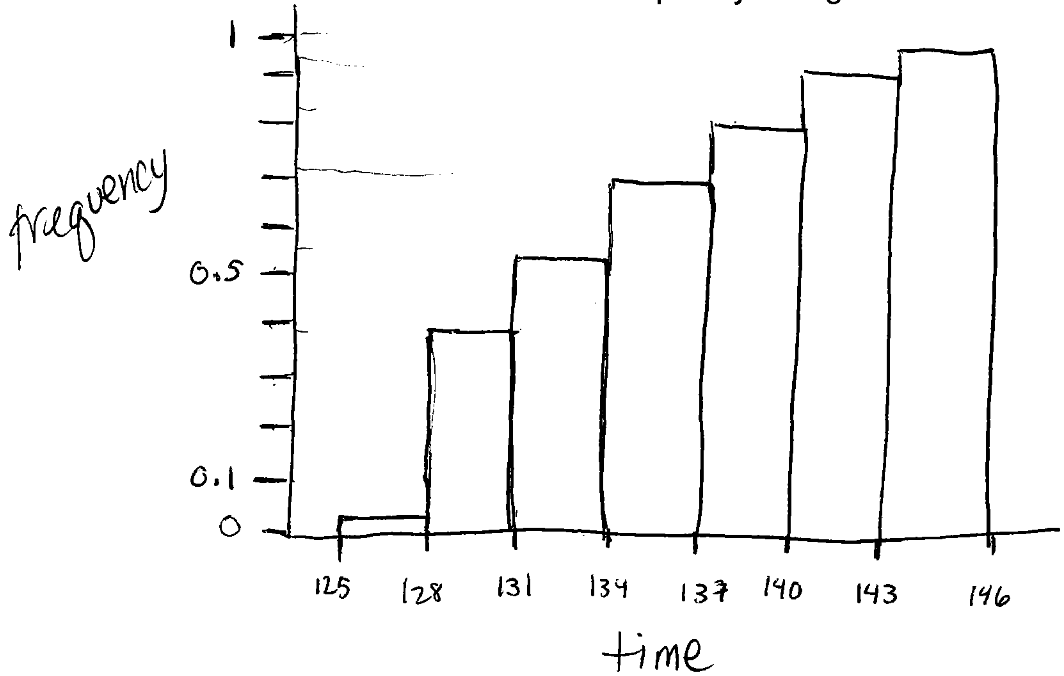
Times:									
143	139	136	139	130	129	129	132	131	129
141	140	142	136	140	132	131	129	128	129
144	137	134	136	135	129	134	129	130	131
144	137	131	134	130	129	128	128	127	

a. Create a frequency histogram. The table below may help you with parts (a) and (b).

Boundaries	Frequency	Cumulative Frequency	Relative Frequency	Cumulative Relative Frequency
125-128	1	1	0.0256	0.0256
128-131	14	15	0.358	0.3846
131-134	6	21	0.154	0.5386
134-137	7	28	0.179	0.7176
137-140	4	32	0.103	0.8206
140-143	4	36	0.103	0.9236
143-146	3	39	0.077	1.0006
total	39	—	1.0006 ≈ 1	—



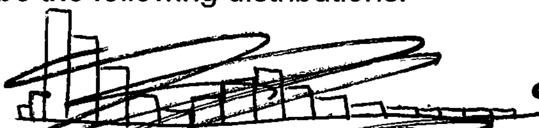
b. Create a cumulative relative frequency histogram



* shape, center, spread!

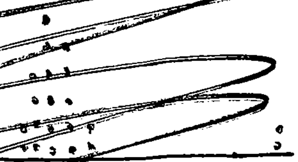
6. Describe the following distributions:

   a.



~~right skewed~~ ~~2 peaks~~ ~~unimodal~~

   b.



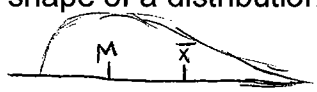
~~roughly symmetric~~ ~~clustered~~ ~~2 poss. outliers~~

   c.

~~left skewed~~ ~~granularity~~

See calculator pics

7. Describe the shape of a distribution that has a mean of 170 and a median of 120.



right skewed

Use the following set of data for numbers 8-10.

weights of 18 year old males											
130	201	190	234	188	160	162	120	124	199	142	179
128	158	202	280	189	172	135	110	187	200	165	188
145	167	122	151	187	174	162	132	195	132	137	186
188	166	204	211	164	145	184	124	192	160	175	180

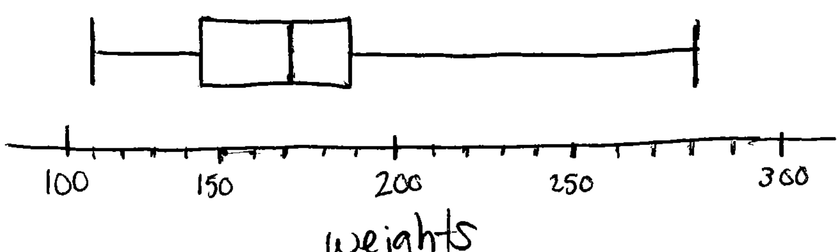
$n = 48$

8. Find the mean and standard deviation of the data.

   $\bar{x} = 169.29$ $s = 32.937$

9. Find the 5 number summary for the data and represent it graphically.

   min = 110
   $Q_1 = 143.5$
   M = 169.5
   $Q_3 = 188.5$
   max = 280



10. Test for possible outliers. Are there any in your opinion (based on the data and the test)?

   IQR = 45
   $45 \times 1.5 = 67.5$
   $Q_1 - 67.5 = 76$
   $Q_3 + 67.5 = 256$

   * anything below 76 & above 256

   280 — yes - look @ histogram & IQR outlier test

11. Which measures of center and spread would you use to represent this data? Why?

   5 # summary - histogram is skewed rt.

12. I have a distribution with the following summary statistics:

mean =	58	Q1 =	48
st. dev. =	12	Q3 =	70
Median =	50	IQR =	22

a. I have now decided that I want to divide each observation by 15. What will the following statistics be for my new set of data?

mean =	3.867	Q1 =	3.2
st. dev. =	0.8	Q3 =	4.667
Median =	3.33	IQR =	1.467

b. I have now decided that I want to go back to my original data set and add 25 to each observation. What will the following statistics be for my new set of data?

mean =	83	Q1 =	73
st. dev. =	12	Q3 =	95
Median =	75	IQR =	22

13. Which of the following summary statistics are resistant to outliers?

~~Mean~~ ~~Standard Deviation~~
Median Quartiles
IQR ~~Variance~~

14. What is the area under a density curve? 1

15. The (mean or median) of a density curve is the equal-areas point, the point that divides the area under the curve in half.

16. The (mean or median) of a density curve is the balance point, at which the curve would balance if made of solid material.

17. If a density curve is skewed to the right, the (mean or median) will be further to the right than the (mean or median).

18. What is the difference between x-bar and μ?

for sample

for population

19. What is the difference between s and σ?

20. How many normal distributions are there?

infinite

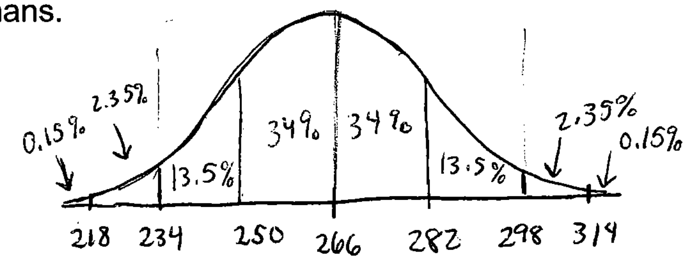
21. How many Standard Normal Distributions are there?

only 1

22. What is the mean and standard deviation of the Standard Normal Distribution?

$N(0,1)$

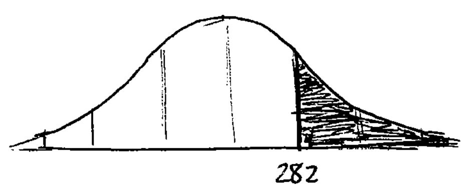
23. Sketch the graph of N(266, 16), the distribution of pregnancy length from conception to birth for humans.



Use the empirical rule (the 68-95-99.7 rule) for problems 24-28.

24. Find the length of the longest 16% of all pregnancies. Sketch and shade a normal curve for this situation.

$X \geq 282$



25. Find the length of the middle 99.7% of all pregnancies.

$\mu - 2\sigma$ $\mu + 2\sigma$

$218 \leq X \leq 314$

26. Find the length of the shortest 2.5% of all pregnancies.

$\mu - 2\sigma$

$X \leq 234$

27. What percentile rank is a pregnancy of 218 days?

0.13th percentile

$P(x<218) = \text{normalcdf}(-E99, 218, 266, 16)$

28. What percentile rank is a pregnancy of 298 days?

97.7th percentile

$P(x<298) = \text{normalcdf}(-E99, 298, 266, 16)$

29. What z-score does a pregnancy of 257 days have?

$$z = \frac{257-266}{16} = -0.5625$$

30. What percent of humans have a pregnancy lasting less than 257 days?

$P(x<257) = \text{normalcdf}(-E99, 257, 266, 16)$
$= 0.28689 \approx 28.689\%$

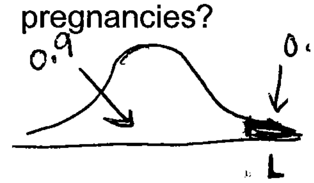
31. What percent of humans have a pregnancy lasting longer than 280 days?

$P(x>280) = \text{normalcdf}(280, E99, 266, 16) = 0.19079$
$\approx 19.08\%$

32. What percent of humans have a pregnancy lasting between 260 and 270 days?

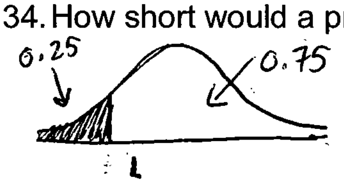
$P(260<x<270) = \text{normalcdf}(260, 270, 266, 16) = 0.24488$
$\approx 24.5\%$

33. How long would a pregnancy have to last to be in the longest 10% of all pregnancies?



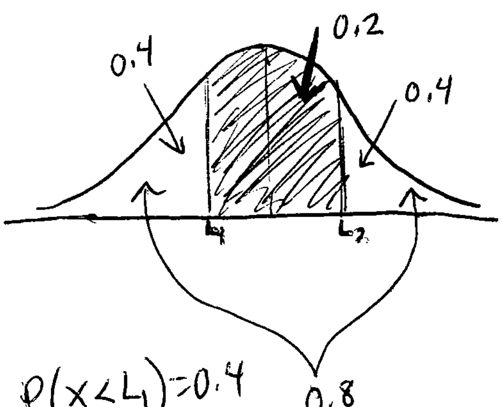
$P(x<L) = 0.9$
$L = \text{invnorm}(0.9, 266, 16) \approx 286.5$

34. How short would a pregnancy be to be in the shortest 25% of all pregnancies?



$P(x<L) = 0.25$
$L = \text{invnorm}(0.25, 266, 16) \approx 255.21$

35. How long would a pregnancy be to be in the middle 20% of all pregnancies?



$L_1 = \text{invnorm}(0.4, 266, 16) \approx 261.95$
$L_2 = \text{invnorm}(0.6, 266, 16) \approx 270.05$

$261.95 < x < 270.05$

$P(x<L_1) = 0.4$
$P(x<L_2) = 0.6$

$N(1500, 75)$

36. The life expectancy of a particular brand of light bulb is normally distributed with a mean of 1500 hours and a standard deviation of 75 hours.

a. What is the probability that a light bulb will last less than 1410 hours?

$P(X<1410) = \text{normalcdf}(-E99, 1410, 1500, 75) = 0.115$

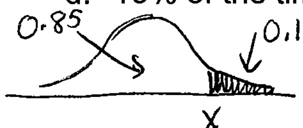
b. What is the probability that a light bulb will last more than 1550 hours?

$P(X>1550) = \text{normalcdf}(1550, E99, 1500, 75) = 0.2525$

c. What is the probability that a light bulb will last between 1563 and 1648 hours?

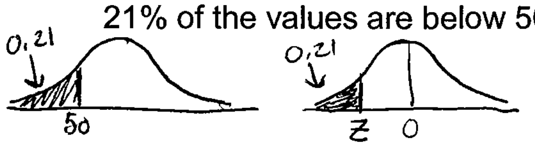
$P(1563<X<1648) = \text{normalcdf}(1563, 1648, 1500, 75) = 0.1762$

d. 15% of the time a light bulb will last more than how many hours?



$\text{invnorm}(0.85, 1500, 75) = 1577.73$

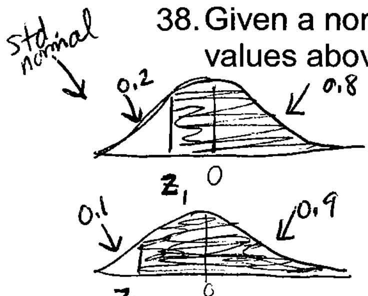
37. Given a normal distribution with a standard deviation of 10, what is the mean (μ) if 21% of the values are below 50?



$\text{invnorm}(0.21, 0, 1) = -0.8064$

$-0.8064 = \frac{50-\mu}{10} \Rightarrow \mu = 58.064$

38. Given a normal distribution with 80% of the values above 125 and 90% of the values above 110, what are the mean and standard deviation?



$z_1 = \text{invnorm}(0.2, 0, 1) = -0.8416 = \frac{125-\mu}{\sigma}$, $\sigma = \frac{125-\mu}{-0.8416}$

$z_2 = \text{invnorm}(0.1, 0, 1) = -1.2816 = \frac{110-\mu}{\sigma}$, $\sigma = \frac{110-\mu}{-1.2816}$

$\frac{125-\mu}{-0.8416} = \frac{110-\mu}{-1.2816}$

$\mu = 153.69$

$\sigma = 34.08$

39. A water fountain is designed to dispense a volume of 12.2 oz. with a standard deviation of 0.5 oz. assume a normal distribution $N(12.2, 0.5)$

a. What percentage of cups end up with at least 12 oz.?

$P(X \geq 12) = \text{normalcdf}(12, E99, 12.2, 0.5) = 0.6554 = 65.54\%$

b. 75% of the cups contain more than how much water?

$P(X<L) = 0.25$ $\text{invnorm}(0.25, 12.2, 0.5) = 11.863$

c. Find the IQR ~~for the~~ amount of water dispensed.

$Q_1 = \text{invnorm}(0.25, 12.2, 0.5) = 11.863$

$Q_3 = \text{invnorm}(0.75, 12.2, 0.5) = 12.537$

$12.537 - 11.863 = 0.6742$

d. Find the 90th percentile for the amount of water dispensed.

the obs that has 90% of the obs below it

$X = \text{invnorm}(0.9, 12.2, 0.5)$

$P(X<L) = 0.9$